Lecture 8. Models for Count Response

Nan Ye

School of Mathematics and Physics
University of Queensland

Examples of Count Responses

Traffic modelling

Predict the number of vehicles going from one place to another.

Behavior modelling

Predict the number of days absent from school.

Mineral exploration

Predict number of occurrences of mineral deposits at different locations.

Manufacturing

Predict number of wave damage incidents to ships.

This Lecture

- Model choices
- Poisson regression
- Overdispersion
- Quasi-Poisson regression
- Negative binomial regression

Models for Count Responses

Structure

- The response function need to be non-negative
  - The log link $g(\mu) = \ln \mu$ is often used.
  - The identity link $g(\mu) = \mu$ is sometimes used (with care).
- The exponential family need to be a distribution on counts

  Poisson distribution, negative binomial distribution (with fixed r)

Poisson Regression

Recall

- When Y is a count, we can use exponentiation to map $\beta^\top \mathbf{x}$ to a non-negative value, and use the Poisson distribution to model $Y \mid \mathbf{x}$, as follows.

$$\begin{aligned} \text{(systematic)} \quad & \mathbb{E}(Y \mid \mathbf{x}) = \exp(\beta^\top \mathbf{x}). \\ \text{(random)} \quad & Y \mid \mathbf{x} \text{ is Poisson distributed.} \end{aligned}$$

- Or more compactly,

$$Y \mid \mathbf{x} \sim Po\left(\exp(\beta^\top \mathbf{x})\right),$$

where $Po(\lambda)$ is a Poisson distribution with parameter λ.

- The Poisson regression model can be explicitly written as

$$p(y \mid \mathbf{x}, \beta) = \frac{\exp(y\beta^\top \mathbf{x})}{y!} \exp(-e^{\beta^\top \mathbf{x}}).$$

- Given $\mathbf{x}$, we can predict Y as the mode

$$\arg\max_y p(y \mid \mathbf{x}, \beta) = \lfloor \exp(\beta^\top \mathbf{x}) \rfloor, \lceil \exp(\beta^\top \mathbf{x}) \rceil - 1.$$

Parameter interpretation

- $\mu = \exp(\beta^\top \mathbf{x})$.
- One unit increase in x_i changes the mean by a factor of e^{β_i}.

Fisher scoring

- Let $\mu_i = \exp(\mathbf{x}_i^\top \beta)$.
- Then the gradient and the Fisher information are

$$\nabla \ell(\beta) = \sum_i (y_i - \mu_i)\mathbf{x}_i,$$
$$I(\beta) = \sum_i \mu_i \mathbf{x}_i^\top \mathbf{x}_i,$$

- Fisher scoring updates β to

$$\beta' = \beta + I(\beta)^{-1} \nabla \ell(\beta).$$

- Let $\mathbf{X}$ be the design matrix, and

$$\boldsymbol{\mu} = (\mu_1, \dots, \mu_n),$$
$$W = \mathrm{diag}\left(\mu_1, \dots, \mu_n\right).$$

- In matrix notation, the gradient and the Fisher information are

$$\nabla \ell(\beta) = \mathbf{X}^\top(\mathbf{y} - \boldsymbol{\mu}),$$
$$I(\beta) = \mathbf{X}^\top W \mathbf{X},$$

Example

Data

```
> library(MASS) # contains the quine dataset
> dim(quine)
[1] 146   5
> head(quine)
  Eth Sex Age Lrn Days
1   A   M  F0  SL    2
2   A   M  F0  SL   11
3   A   M  F0  SL   14
4   A   M  F0  AL    5
5   A   M  F0  AL    5
6   A   M  F0  AL   13
```

- Subjects are 146 children from Walgett, New South Wales, Australia.
- The Culture, Age, Sex and Learner status and the number of days absent from school in a particular school year were recorded.
- Type `help(quine)` to read more about the dataset.

Poisson regression

```
> fit.po <- glm(Days ~ Sex + Age + Eth + Lrn, data=quine,
    family=poisson)
> summary(fit.po)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.71538    0.06468  41.980  < 2e-16 ***
SexM         0.16160    0.04253   3.799 0.000145 ***
AgeF1       -0.33390    0.07009  -4.764 1.90e-06 ***
AgeF2        0.25783    0.06242   4.131 3.62e-05 ***
AgeF3        0.42769    0.06769   6.319 2.64e-10 ***
EthN        -0.53360    0.04188 -12.740  < 2e-16 ***
LrnSL        0.34894    0.05204   6.705 2.02e-11 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)
```

First thought...

- All covariates are highly significant according to Wald's test.
- Looks like we have a very good model!

Recall

- With a mis-specified model, asymptotic normality still holds, but the mean and the covariance matrix of the asymptotic distribution now depend on both the model class and the *unknown* true distribution.
- The confidence interval and the distribution of Wald's statistics cannot be computed, and can only be applied (*with caution*) if the model is not too much away from reality.

Are we sure that the model is well-specified?

Predictive performance on training set

```
> mean(quine$Days)
[1] 16.4589
> mean(abs(quine$Days - predict(fit.po, type='response')))
[1] 11.04622
> summary(quine$Days)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00    5.00   11.00   16.46   22.75   81.00
> summary(predict(fit.qpo, type='response'))
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
6.346  10.821  15.339  16.459  22.984  32.582
```

- Mean absolute error is high ($11.04622/16.4589 \approx 67\%$).
- y_i's have very large range as compared to μ_i's, which is quite unlikely if the data follows a Poisson distribution.
- We are observing *overdispersion*: variance in data is larger than expected based on the model.

Overdispersion for Poisson

Example 1. Clustering

- Consider the clustered Poisson process

$$\begin{aligned} &N \sim Po(\mu), \\ &Y = Z_1 + \ldots + Z_N, \qquad Z_i\text{'s are i.i.d.}, \end{aligned}$$

  Here we think of each Z_i as the count in a cluster.
- The mean and variance of Y are

$$\mathbb{E}(Y) = \mathbb{E}(N)\,\mathbb{E}(Z), \qquad \text{var}(Y) = \mathbb{E}(N)\,\mathbb{E}(Z^2).$$

- If Z_i's take value 1 with probability 1, then $Y \sim Po(\mu)$.
- Relative to Poisson: we observe overdispersion if $\mathbb{E}(Z^2) > \mathbb{E}(Z)$, and underdispersion if $\mathbb{E}(Z^2) < \mathbb{E}(Z)$.

Example 2. Inter-subject variability

- Consider the Gamma mixture of Poisson distributions

$$\begin{aligned} \lambda &\sim \Gamma(\text{mean} = \mu, \text{var} = \mu/\phi), \\ Y &\sim Po(\lambda). \end{aligned}$$

  Here we treat each individual as having different mean λ.

- Y follows a negative binomial distribution

$$Y \mid \mu, \phi \sim NB\left(\text{mean} = \mu, p = \frac{1}{1+\phi}\right).$$

- $\text{var}(Y) = \mu/(1-p) > \mu$, so we have overdispersion relative to Poisson.

Quasi-Poisson Regression

- Quasi-Poisson regression model introduces an additional dispersion paramemeter ϕ.
- It replaces the original model variance V_i on $\mathbf{x}_i$ by ϕV_i.
- $\phi > 1$ is used to accommodate overdispersion relative to the original model.
- $\phi < 1$ is used to accommodate underdispersion relative to the original model.
- ϕ is usually estimated separately after estimating β.

Quasi-Poisson regression

```
> fit.qpo <- glm(Days ~ Sex + Age + Eth + Lrn, data=quine,
    family=quasipoisson)
> summary(fit.qpo)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.7154     0.2347  11.569  < 2e-16 ***
SexM          0.1616     0.1543   1.047 0.296914
AgeF1        -0.3339     0.2543  -1.313 0.191413
AgeF2         0.2578     0.2265   1.138 0.256938
AgeF3         0.4277     0.2456   1.741 0.083831 .
EthN         -0.5336     0.1520  -3.511 0.000602 ***
LrnSL         0.3489     0.1888   1.848 0.066760 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for quasipoisson family taken to be 13.16691)
```

- Estimated coefficients of Poisson regression and quasi Poisson regression are the same (though printed differently).
- The dispersion parameter for quasi Poisson is 13.16691, indicating overdispersion relative to Poisson.
- Quasi Poisson indicates that only Ethnicity and intercept are significant.

Negative Binomial Regression

- Uses the negative binomial distribution as the random component.
- This is not a GLM (unless we fixed the r parameter in $NB(r, p)$).
- The parameters can still be estimated using MLE.

Using glm.nb from the MASS library

```
> fit.nb <- glm.nb(Days ~ Sex + Age + Eth + Lrn, data=quine)
> summary(fit.nb)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.89458    0.22842  12.672  < 2e-16 ***
SexM         0.08232    0.15992   0.515 0.606710
AgeF1       -0.44843    0.23975  -1.870 0.061425 .
AgeF2        0.08808    0.23619   0.373 0.709211
AgeF3        0.35690    0.24832   1.437 0.150651
EthN        -0.56937    0.15333  -3.713 0.000205 ***
LrnSL        0.29211    0.18647   1.566 0.117236
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for Negative Binomial(1.2749) family taken to
    be 1)
```

We get roughly the same qualitative conclusion as quasi Poisson.

Dunning-Kruger Effect
in statistics...

A very wrong model can be very confident.

Validate model assumptions before you trust.

What You Need to Know

- Model choices
- Poisson regression: $p(y \mid \mathbf{x}, \beta)$, parameter interpretation, Fisher scoring, Dunning-Kruger effect.
- Understand how overdispersion can occur relative to Poisson.
- Using quasi-Poisson regression to model data with variance different from mean.
- Using negative binomial regression to model data with variance larger than mean.